

ANALISIS STATISTIK TINGKAT PENGANGGURAN TERBUKA DI JAWA TIMUR TAHUN 2012

Nama Mahasiswa : Nugroho Manggala Putra
NRP : 1310 100 019
Jurusan : Statistika
Dosen Pembimbing : Dr. Dra. Ismaini Zain, M.Si

ABSTRAK

Pengangguran merupakan masalah yang pasti dihadapi oleh setiap wilayah di Indonesia. Jawa timur sebagai provinsi terluas diantara provinsi di Pulau Jawa dengan jumlah penduduk terpadat di Indonesia dan terus mengalami pertumbuhan setiap tahun juga menghadapi masalah tersebut. Menurut BPS dalam Dinas Tenaga Kerja (2012) jumlah pengangguran mencapai 819.563 jiwa tercatat pada bulan Agustus tahun 2012. Pendekatan *principal component regression* dan regresi ridge dipilih karena beberapa penelitian sebelumnya menyatakan bahwa metode ini dapat menangani kasus multikolinieritas. *K-means cluster* merupakan salah satu prosedur pengelompokan dengan menempatkan obyek berdasarkan *means* kelompok terdekat. Analisis diskriminan berguna untuk membentuk sebuah model prediktif dari beberapa kelompok (*group*) berdasarkan pada karakteristik masing-masing variabel. Pemodelan menggunakan PCA menghasilkan model yang tidak cukup bagus untuk menggambarkan tingkat pengangguran terbuka secara actual, sehingga digunakan regresi ridge untuk mendapatkan model yang lebih robust. Pengelompokan dengan *k-means* memiliki karakteristik unik yang berbeda antara satu kelompok dengan lainnya. Nilai *Apperent Error Rates* (APER) klasifikasi kabupaten/kota di Jawa Timur berdasarkan faktor-faktor yang mempengaruhi tingkat pengangguran terbuka sebesar 0,132 yang berarti bahwa akurasi pada model mencapai 86,8 persen.

Kata Kunci: *APER, K-means Cluster, PCA, Ridge Regression, TPT*

STATISTICAL ANALYSIS OF THE OPEN UNEMPLOYMENT RATE IN EAST JAVA

Nama Mahasiswa : Nugroho Manggala Putra
NRP : 1310 100 019
Jurusan : Statistika
Dosen Pembimbing : Dr. Dra. Ismaini Zain, M.Si

ABSTRAK

Unemployment is one of the biggest problems experienced by Indonesia. East Java is the largest province than any province in Java. In August 2012, the unemployment is 819.563 people. Principal component regression (PCR) and ridge regression is used in this experiment because this method can handle the multicollinearity. K-means cluster is one of methods to classify object base on the closest means from the group. Discriminant analysis can make a model from groups base on characteristic of variables. The model in principal component regression is not good enough to describe the actual problem of open unemployment, so ridge regression is used to obtain a robust model. Classifying with k-means has unieque characteristics that differ from one group to another. The Apperent Error Rates (APER) of distrcts/cities in East Java base on factors that affect the open unemployment rate at 0,132. It means that the accuracy of the models reached 86,6 percent.

Kata Kunci: *APER, K-means Cluster, PCA, Ridge Regression, Open Unemployment*

BAB II TINJAUAN PUSTAKA

2.1 Regresi Linier

Metode regresi adalah metode yang digunakan untuk menyatakan pola hubungan antara variabel respon (Y) dan variabel prediktor (X). Pengamatan sebanyak n dengan variabel prediktor (X) sebanyak p maka model regresi tersebut dapat ditulis sebagai berikut.

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i$$

dengan $i = 1, 2, \dots, n$

Dengan β_0 sebagai parameter konstan, β_k adalah parameter model untuk $k=1, 2, \dots, p$ dan ε merupakan error yang diasumsikan independen, identik dan berdistribusi normal (Draper dan Smith, 1992). Persamaan regresi tersebut dapat dinotasikan dalam bentuk matrik sebagai berikut.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

dengan:

$$\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_n \end{bmatrix}^T$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

Parameter model regresi diduga menggunakan metode OLS (*Ordinary Least Square*) yaitu dengan meminimumkan jumlah kuadrat *error*nya. Sedangkan, penaksir parameter model didapat dari persamaan (Draper dan Smith, 1992):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Tabel 2.1 Tabel ANOVA

Sumber	Db/df	SS	MS	F _{hitung}
Regresi	p	SSR	SSR/k	MSR/MSE
Residual Error	n-p-1	SSE	SSE/(n-k-1)	
Total	n-1	SST		

Keterangan:

$$SST = \mathbf{Y}^T \mathbf{Y} - n\bar{\mathbf{Y}}^2$$

$$SSR = \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} - n\bar{\mathbf{Y}}^2$$

$$SSE = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y}$$

Tabel 2.1 menjelaskan tentang tabel ANOVA secara ringkas. Sedangkan koefisien determinasi (R^2) merupakan nilai yang menjelaskan tingkat hubungan variabel independen pada model regresi.

$$R^2 = \frac{SSR}{SST}$$

Koefisien regresi diuji secara serentak untuk mengetahui apakah secara bersama-sama mempunyai pengaruh yang signifikan terhadap model. Hipotesis dalam pengujian ini adalah

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1: \text{Paling sedikit ada satu } \beta_k \neq 0, k = 1, 2, \dots, p$$

Statistik uji yang digunakan adalah

$$F_{\text{hit}} = \frac{MSR}{MSE} = \frac{SSR/p}{SSE/(n-p-1)}$$

Apabila $F_{\text{hit}} > F_{(\alpha; p; n-p-1)}$ atau $P_value < \alpha$, maka H_0 ditolak artinya paling sedikit ada satu β_k yang tidak sama dengan nol atau paling sedikit ada satu dari variabel bebas yang memiliki pengaruh signifikan terhadap variabel respon.

Pengujian selanjutnya yang dilakukan adalah pengujian signifikansi parameter secara parsial. Hipotesis pengujian signifikansi parameter secara parsial adalah sebagai berikut.

$$H_0: \beta_k = 0$$

$$H_1: \beta_k \neq 0$$

Statistik uji yang digunakan adalah

$$t_{\text{hit}} = \frac{\hat{\beta}_k}{s(\hat{\beta}_k)} = \frac{b_k}{s(b_k)}$$

Dimana:

$$s(\hat{\beta}_k) = \text{standar error koefisien} = (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2$$

2.1.1 Deteksi *Outlier*

Outlier atau yang biasa disebut dengan istilah pencilan merupakan observasi dengan nilai yang lebih besar dari nilai respon maupun nilai prediktor pada umumnya. Salah satu langkah untuk mengidentifikasi *outlier* ini adalah dengan menggunakan *differences in FITS* (DFFITS) yang diperkenalkan oleh Welsch dan Kuh (1977). Prinsip dasar metode ini adalah menghitung perbedaan antara nilai taksiran, dengan atau tidak menggunakan observasi ke-*i*. Mengevaluasi observasi dengan menggunakan metode ini yakni dengan

$$\text{DFITS} = \frac{\hat{y}_i - \hat{y}_{i(i)}}{\sqrt{\text{MSE}_{i(i)} h_i}} = e_i \sqrt{\frac{n - p - 1}{\text{SSE}(1 - h_i) - e_i^2}} \left[\sqrt{\frac{h_i}{1 - h_i}} \right]$$

Dimana:

e_i : residual ke-*i*

h_i : elemen diagonal ke-*i* dari $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

\hat{y}_i : taksiran respon ke-*i*

$\hat{y}_{i(i)}$: nilai taksiran yang dihitung tanpa observasi ke-*i*

$\text{MSE}_{i(i)}$: *mean square error* tanpa observasi ke-*i*

p merupakan banyak prediktor (termasuk β_0) dan *n* merupakan

banyak observasi. Apabila didapatkan nilai $|\text{DFITS}| > 2\sqrt{p/n}$ maka observasi ke-*i* merupakan *outlier* yang perlu ditangani, namun

sebaliknya bila nilai $|\text{DFITS}| < 2\sqrt{p/n}$ maka *outlier* tersebut tidak perlu dihilangkan dari pengamatan (Belsley, Kuh dan Welsch, 1990).

2.2 Uji Multikolinieritas

Salah satu syarat yang harus dipenuhi dalam regresi dengan menggunakan beberapa variabel prediktor adalah tidak adanya korelasi antar variabel prediktor. Adanya korelasi antar variabel prediktor menyebabkan taksiran parameter regresi yang dihasilkan akan memiliki error yang sangat besar. Pendeteksian kasus kolinieritas/multikolinieritas dapat dilihat melalui nilai VIF

(Variance Inflation Factors). Variance Inflation Factors (VIF) dinyatakan dalam (Hocking, 1996)

$H_0 : \rho = 1$ (tidak terdapat multikolinieritas)

$H_1 : \rho \neq 1$ (terdapat multikolinieritas)

Statistik uji yang digunakan adalah

$$VIF = 1/(1 - R_i^2)$$

VIF lebih besar dari 10 menunjukkan adanya multikolinieritas antar variabel-variabel prediktor.

2.3 Pemeriksaan Asumsi Residual Identik, Independen dan Berdistribusi Normal (IIDN)

Setelah melakukan pengujian signifikansi parameter, dilanjutkan dengan pengujian terhadap asumsi residual yang identik, independen dan berdistribusi normal.

2.3.1 Asumsi Residual Identik

Salah satu asumsi yang penting model regresi adalah varians residual bersifat homoskedastisitas yaitu varians residual bersifat identik. Cara mendeteksi heteroskedastisitas adalah melalui uji *glejser* yang dilakukan dengan cara meregresikan harga mutlak residual $|e|$ dengan variabel prediktor (x). Model

regresi tersebut dinyatakan dengan $|e_i| = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i$. Jika

diketahui tidak terdapat variabel prediktor yang signifikan dalam model, maka hal tersebut mengindikasikan bahwa residual cenderung identik. Dan dapat dinyatakan pada hipotesis sebagai berikut.

$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$ (Residual Identik)

$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2$ (Residual Tidak Identik)

Dimana $i = 1, 2, \dots, n$

$$F_{\text{hit}} = \frac{(|e|^T X^T |e| - n|\bar{e}|^2)/p}{(|e|^T e - \hat{\beta} X^T |e|)/(n - p - 1)}$$

Dimana $|e|$ merupakan matrik dengan harga mutlak dan $\hat{\beta}$ adalah matriks koefisien estimasi. Pengambilan keputusan adalah apabila $F_{\text{hit}} > F_{\alpha (p, n-p-1)}$ maka H_0 ditolak pada tingkat signifikansi

α , artinya paling sedikit ada satu $\sigma_i^2 \neq \sigma^2$ atau tidak terjadi homoskedastisitas (tidak identik).

2.3.2 Asumsi Residual Independen

Uji independen atau uji autokorelasi residual untuk mengetahui apakah ada korelasi antara residual. Salah satu metode yang dapat digunakan untuk menguji independensi tersebut adalah plot ACF (*Autocorrelation Function*), dengan hipotesis dan statistic uji sebagai berikut.

$$H_0 : \rho = 0 \text{ (Residual Independen)}$$

$$H_1 : \rho \neq 0 \text{ (Residual Tidak Independen)}$$

$$\rho_k = \frac{\sum_{t=k+1}^n (e_{t-k} - \bar{e})(e_t - \bar{e})}{\sum_{t=1}^n (e_t - \bar{e})^2}$$

Nilai autokorelasi yang keluar dari batas signifikansi mengindikasikan asumsi residual independen tidak terpenuhi dan begitu pula sebaliknya.

2.3.3 Asumsi Residual Berdistribusi Normal

Pengujian asumsi residual Normal ($0, \sigma^2$) dapat dilakukan melalui uji Kolmogorov Smirnov. Hipotesis dan statistik uji yang digunakan adalah

$$H_0 : F_0(e) = F(e) \text{ (Residual berdistribusi Normal } (0, \sigma^2))$$

$$H_1 : F_0(e) \neq F(e) \text{ (Residual tidak berdistribusi Normal } (0, \sigma^2))$$

Statistik uji

$$D = \max |F_0(e) - S(e)|$$

Dimana $F(e)$ adalah fungsi distribusi kumulatif distribusi normal dari residual, $F_0(e)$ adalah fungsi distribusi kumulatif teoritis dari residual, sedangkan $S(e) = i/n$ merupakan fungsi peluang kumulatif pengamatan dari suatu sampel random dengan i adalah pengamatan dan n adalah jumlah pengamatan.

Pengambilan keputusan adalah H_0 ditolak jika $|D| > q_{(1-\alpha)}$ dimana q adalah nilai berdasarkan tabel Kolmogorov Smirnov.

2.4 Principal Component Regression (PCR)

Principle Component Regression atau Regresi Komponen Utama merupakan salah satu metode yang biasa digunakan untuk mengatasi kasus multikolinieritas. Prinsip dasar metode regresi ini adalah memasukkan seluruh variabel independen dan mengakomodasi adanya multikolinieritas antar variabel independen dengan cara mengelompokkan variabel yang saling berkorelasi cukup tinggi dalam sebuah variabel baru. Sehingga mereduksi banyaknya dimensi regresi dan antar variabel baru tersebut tidak saling berkorelasi.

Langkah-langkah pada PCR terbagi menjadi beberapa tahapan. Melakukan standarisasi/pembakuan data variabel independent x_j merupakan langkah pertama yang harus dilakukan.

$$z_j = \frac{x_j - \bar{x}_j}{s_j}$$

Matrik korelasi antar variabel independen berguna untuk melihat seberapa tinggi hubungan antar setiap variabel independen. Setelah mengetahui hubungan antar variabel independen, membangkitkan variabel baru adalah langkah yang harus ditempuh untuk mendapatkan variabel baru yang saling *independent*.

$$PC_1 = a_{11}Z_1 + a_{12}Z_2 + \dots + a_{1k}Z_k$$

$$PC_2 = a_{21}Z_1 + a_{22}Z_2 + \dots + a_{2k}Z_k$$

.....

$$PC_k = a_{k1}Z_1 + a_{k2}Z_2 + \dots + a_{kk}Z_k$$

atau bisa dituliskan dalam bentuk $PC_j = \mathbf{a}_j \mathbf{T} \mathbf{z}$, nilai \mathbf{a} adalah *eigen-vector* dari *eigen value* ke- j dari matriks korelasi antar variabel *independent*. Banyaknya PC ditentukan berdasarkan kriteria *eigen value* ≥ 1 . Meregresikan variabel respon y dengan skor PC hingga menyatakan model regresi Y dengan PC ke dalam model Y dengan z hingga ke dalam x .

2.5 Regresi Ridge

Regresi Ridge pertama kali dikemukakan oleh A. E. Hoerl pada 1962. Metode ini digunakan untuk mengatasi kondisi buruk yang diakibatkan oleh korelasi yang tinggi antara beberapa variabel prediktor di dalam model (multikolinieritas), sehingga

menyebabkan nilai dugaan parameter model yang tidak stabil (Draper dan Smith, 1992).

$$z_j = \frac{x_j - \bar{x}_j}{s_j \sqrt{n-1}}$$

Regresi ridge menghasilkan nilai dugaan yang bersifat bias tetapi efisien (mempunyai varians yang lebih kecil bila dibandingkan dengan varians penduga regresi linier berganda). Metode kuadrat terkecil digunakan untuk menduga parameter regresi ridge dengan menambahkan kendala tertentu pada nilai diagonal matriks $\mathbf{X}^T \mathbf{X}$. besarnya kendala yang ditambahkan mencerminkan besarnya bias penduga pada regresi ridge. Besarnya kendala tersebut selanjutnya disebut dengan d .

Misalkan \mathbf{Z} adalah matrik \mathbf{X} yang telah dipusatkan dan diskalakan, sehingga matriks korelasi $\mathbf{Z}^T \mathbf{Z}$ dan nilai eigen matriks korelasi adalah r . Nilai dugaan β^r untuk variabel bebas Z_k dapat diketahui melalui

$$\beta^r = (\mathbf{Z}^T \mathbf{Z} + d\mathbf{I})^{-1} \mathbf{Z}^T \mathbf{Y}$$

dengan nilai d yang terletak dalam selang $0 < d < 1$ (Draper dan Smith, 1992). Varians dari regresi ridge adalah

$$s(\beta^r) = \sigma^2 [d\mathbf{I} + (\mathbf{Z}^T \mathbf{Z})]^{-1} \mathbf{Z}^T \mathbf{Z} [d\mathbf{I} + (\mathbf{Z}^T \mathbf{Z})]^{-1}$$

Estimator regresi ridge dengan berbagai nilai d digambarkan dalam bentuk plot secara bersama-sama yang disebut dengan *Ridge Trace*. Pemilihan besarnya d perlu diperhatikan sebab tetapan yang diinginkan adalah tetapan yang menghasilkan bias kecil dan menghasilkan penduga yang relatif stabil.

2.6 Analisis Klaster

Metode pengelompokan hirarki digunakan apabila belum ada informasi jumlah kelompok. Sedangkan metode pengelompokan bukan hirarki bertujuan mengelompokkan n obyek ke dalam k kelompok ($k < n$). Salah satu prosedur pengelompokan bukan hirari adalah dengan menggunakan *k-means* yang diperkenalkan oleh James B. Mac Queen pada 1967. Dasar pengelompokan metode ini adalah menempatkan obyek berdasarkan *means* kelompok terdekat. Sehingga metode ini

bertujuan untuk meminimumkan *error* akibat partisi n obyek ke dalam k kelompok. Prinsip dasar metode *k-means* adalah meminimumkan jumlah kuadrat *error* seluruh i kelompok, yaitu

$$SSE = \sum_{i=1}^p \sum_{x_i \in c_i} (c_i - x_i)^2$$

K-means dimulai dengan pemilihan secara acak k dimana nilai k merupakan banyaknya kelompok yang ingin dibentuk. Kemudian nilai-nilai k ditetapkan secara random, untuk sementara nilai tersebut menjadi pusat dari kelompok atau biasa disebut *means*. Selanjutnya jarak setiap data yang ada terhadap masing-masing *means* dihitung menggunakan ukuran ketakmiripan hingga ditemukan jarak yang paling dekat dari setiap data dengan pusat dari kelompok. Setiap data diklasifikasikan berdasarkan kedekatannya dengan pusat dari kelompok. Langkah-langkah tersebut terus dilakukan hingga nilai pusat kelompok tidak berubah (stabil). Adapun rumus untuk menghitung pusat kelompok adalah sebagai berikut.

$$c_{(A)i} = \frac{1}{n_i} \sum_{j=1}^n y_{ij}$$

Dimana

$c_{(A)i}$ = pusat kelompok A pada variabel ke- i

A = 1, 2, ..., p

n_i = jumlah obyek pada variabel ke- i

y_{ij} = nilai dari obyek ke- j pada variabel ke- i

i = 1, 2, ..., p

p = banyak variabel

2.7 Analisis Diskriminan

Analisis diskriminan berguna untuk membentuk sebuah model prediktif dari beberapa kelompok (group) berdasarkan pada karakteristik masing-masing kasus. Prosedur pembentukan fungsi diskriminan berdasarkan pada kombinasi linear dari variabel-variabel prediktor yang memberikan pembeda terbaik dari kelompok-kelompok tersebut. Fungsi dibentuk dari sebuah sampel pada sebuah kasus di dalam sebuah group yang telah diketahui, fungsi lalu dapat diaplikasikan pada kasus baru dengan

pengukuran pada variabel-variabel prediktor yang tidak diketahui masuk pada kelompok mana. Model analisis diskriminan adalah sebagai berikut.

$$Y_i = b_0 + b_1X_{i1} + b_2X_{i2} + b_3X_{i3} + \dots + b_kX_{ik}$$

dimana:

Y : Nilai diskriminan dari responden ke- i . Dengan $i = 1, 2, \dots, k$

X_{ij} : Variabel ke- j dari responden ke- i

b_j : Koefisien diskriminan dari variabel ke- j

Ketepatan data:

$$Y_{cs} = \frac{N_A Y_B + N_B Y_A}{N_A + N_B}$$

$$K = \frac{n}{N} \times 100\%$$

dimana:

Y_A : Nilai diskriminan dari responden ke- i untuk data Y kelompok pertama.

Y_B : Nilai diskriminan dari responden ke- i untuk data Y kelompok kedua.

N_A : Banyaknya sampel kelompok pertama.

N_B : Banyaknya sampel kelompok kedua.

K : Ketepatan data.

n : Banyaknya Y, Y_{cs} yang sama.

N : Banyaknya data total.

Jika $Y > Y_{cs}$, maka baris masuk grup “1”, sedangkan jika $Y < Y_{cs}$, maka baris masuk grup “0”.

2.7.1 Evaluasi Ketetapan Klasifikasi

APER atau *Apparent Error Rates* dapat digunakan untuk mengetahui ketepatan klasifikasi dalam suatu kelompok (Johnson dan Wichern, 2007). Menghitung APER dapat ditunjukkan dengan tabel klasifikasi seperti pada Tabel 2.2.

Tabel 2.2. Klasifikasi Kelompok Sebenarnya dan Kelompok Dugaan

Kelompok Sebenarnya	Kelompok Dugaan	
	π_1	π_2
π_1	n_{11}	n_{12}
π_2	n_{21}	n_{22}

dimana:

n_{11} = Jumlah pengamatan dari π_1 tepat diklasifikasi sebagai π_1

n_{12} = Jumlah pengamatan dari π_1 tepat diklasifikasi sebagai π_2

n_{21} = Jumlah pengamatan dari π_2 tepat diklasifikasi sebagai π_1

n_{22} = Jumlah pengamatan dari π_2 tepat diklasifikasi sebagai π_2

$$APER = \frac{n_{12}+n_{21}}{n_1+n_2} \frac{\text{jumlah total objek yang salah klasifikasi}}{\text{jumlah total sampel}}$$

Pengklasifikasian terhadap jumlah pengamatan atau observasi yang berasal dari dua populasi π_1 dan π_2 dapat diketahui dengan menggunakan aturan dan metode klasifikasi fungsi diskriminan fisher sebagai berikut.

Jika x_0 di alokasikan ke π_1 dengan syarat : $\hat{y}_0 \geq \hat{m}$ atau $\hat{y}_0 \leq \hat{m}$

dimana:

$$\hat{y}_0 = (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} x_0$$

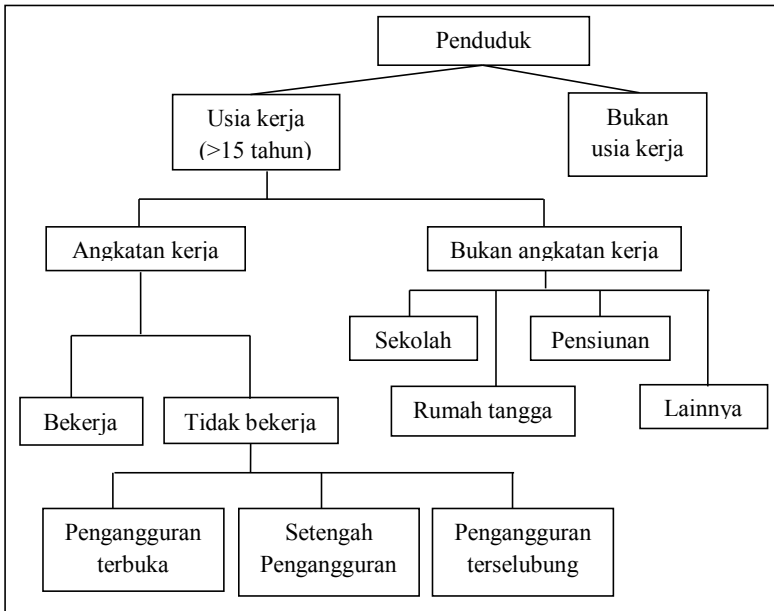
$$\hat{m} = \frac{1}{2} (\bar{x}_1 - \bar{x}_2)' S_{pooled}^{-1} (\bar{x}_1 - \bar{x}_2)$$

sebaliknya x_0 dialokasikan ke π_1 jika $\hat{y}_0 \leq \hat{m}$ atau $\hat{y}_0 - \hat{m} \leq 0$

2.8 Tingkat Pengangguran Terbuka (TPT)

Di Indonesia, angkatan kerja adalah penduduk yang berumur 15 tahun keatas yang secara aktif melakukan kegiatan ekonomis (Mantra, 2000). Angkatan kerja terdiri dari penduduk bekerja, mempunyai pekerjaan tetap tetapi sementara tidak bekerja, dan tidak mempunyai pekerjaan sama sekali tetapi

mencari pekerjaan secara aktif. Sedangkan penduduk yang berumur 15 tahun atau tidak bekerja atau tidak mencari pekerjaan karena sekolah, mengurus rumah tangga, pensiun, atau secara fisik dan mental tidak memungkinkan untuk bekerja tidak termasuk dalam angkatan kerja. Pertambahan angkatan kerja dipengaruhi oleh meningkatnya jumlah penduduk usia kerja (Mantra, 2000).



Gambar 2.1 Struktur Ketenagakerjaan di Indonesia

Pengangguran sering diartikan sebagai angkatan kerja yang belum bekerja atau bekerja tidak optimal, berdasarkan pengertian tersebut pengangguran dibedakan menjadi tiga yaitu pengangguran terbuka pengangguran terselubung, dan setengah menganggur. Pengangguran terbuka merupakan kelompok penduduk usia kerja yang seminggu terakhir (ketika dilakukan survei) tidak bekerja dan sedang mencari pekerjaan, mempersiapkan usaha, tidak mencari pekerjaan karena merasa sulit mendapat pekerjaan, dan sudah mendapat pekerjaan tetapi

belum bekerja. Pengangguran terselubung adalah seorang yang bekerja tetapi penghasilan yang diperoleh tidak mencukupi kebutuhan hidupnya. Sedangkan, setengah pengangguran adalah seorang yang bekerja kurang dari 35 jam per minggu dan masih mencari pekerjaan/bersedia menerima pekerjaan lain. Struktur yang lebih sederhana tentang ketenagakerjaan dapat dilihat pada Gambar 2.1.

2.9 Faktor-faktor yang Mempengaruhi Tingkat Pengangguran Terbuka

Penyebab terjadinya pengangguran terbuka yang utama adalah semakin meningkatnya jumlah angkatan kerja namun peluang untuk mendapatkan kesempatan kerja sangat kecil (Santoso, 2009). Faktor lain yang dapat mempengaruhi tingkat pengangguran yaitu tidak sesuainya keahlian dan ketrampilan yang dimiliki angkatan kerja dengan spesifikasi pekerjaan perusahaan, ketidaksesuaian antara gaji yang ditawarkan perusahaan dengan yang diminta oleh pekerja. Pertambahan angkatan kerja dipengaruhi oleh meningkatnya jumlah penduduk usia kerja (Mantra, 2000). Di lain sisi, perusahaan formal (industrial) menjadi sektor usaha yang menyerap banyak tenaga kerja, sehingga sedikit banyaknya jumlah perusahaan industri mempengaruhi laju perekonomian daerah yang berdampak pada tingkat pengangguran (Dinas Tenaga Kerja, 2007).

BAB III

METODOLOGI PENELITIAN

3.1 Sumber Data

Sumber data yang digunakan pada penelitian ini adalah data sekunder yang didapatkan dari data Survei Angkatan Kerja Nasional (SAKERNAS) dan Jawa Timur dalam Angka 2013 Badan Pusat Statistik (BPS) Provinsi Jawa Timur, namun data yang digunakan hanya data pada tahun 2012.

3.2 Definisi Operasional

Bekerja: kegiatan dengan maksud memperoleh atau membantu memperoleh penghasilan atau keuntungan, selama paling sedikitnya satu jam dalam seminggu selama pencacahan. Bekerja selama satu jam tersebut harus dilakukan berturut-turut dan tidak terputus.

Mencari pekerjaan dibagi menjadi:

1. Mereka yang bekerja, tetapi karena suatu hal masih mencari pekerjaan.
2. Mereka yang dibebaskan tugas dan akan dipanggil kembali tetapi sedang berusaha untuk mendapatkan pekerjaan.
3. Mereka yang belum pernah bekerja dan sedang berusaha mendapatkan pekerjaan.

Tingkat pengangguran terbuka: rasio antara mereka yang sedang mencari pekerjaan dengan jumlah angkatan kerja.

Angkatan kerja: penduduk berumur 15 tahun keatas yang secara aktif melakukan kegiatan ekonomis.

Laju pertumbuhan daerah: besar kecilnya persentase peningkatan barang dan jasa masyarakat menurut sektor produksi suatu daerah.

Tingkat investasi daerah: pengeluaran atau pembelanjaan suatu daerah yang berfungsi sebagai pembentuk modal dalam rangka mencapai berbagai tujuan pembangunan daerah.

Upah minimum: suatu standar minimum yang digunakan oleh para pengusaha atau pelaku industri untuk memberikan upah kepada pekerja di dalam lingkungan usaha atau kerja.

Kondisi hidup layak: standar kebutuhan yang harus dipenuhi oleh seorang pekerja atau buruh lajang untuk dapat hidup layak baik secara fisik, non fisik dan sosial dalam satu bulan.

3.3 Variabel Penelitian

Tabel 3.1 menunjukkan variabel-variabel penelitian yang digunakan dalam menentukan model adalah sebagai berikut.

Tabel 3.1 Variabel Penelitian

Variabel	Ketereangan
Y	Tingkat pengangguran terbuka per Kab/Kota
X1	Persentase Tenaga Kerja Indonesia masing-masing Kab/Kota
X2	Tingkat Upah Minimum Kab/Kota
X3	Persentase angkatan kerja masing-masing Kab/Kota
X4	Persentase Upah Minimum Kab/Kota dari Kondisi Hidup Layak
X5	Rasio banyaknya perusahaan per 100 angkatan kerja masing-masing Kab/Kota
X6	Laju pertumbuhan ekonomi daerah
X7	Persentase angkatan kerja berpendidikan SMA keatas
X8	Tingkat investasi daerah

3.4 Langkah-langkah Analisis Data

Langkah-langkah yang dilakukan untuk mengkaji karakteristik data pada masing-masing kelompok Kabupaten/Kota di Jawa Timuir adalah sebagai berikut.

1. Melakukan analisis deskriptif untuk variabel tingkat pengangguran terbuka masing-masing Kabupaten/Kota.
2. Melakukan analisis deskriptif untuk setiap variabel prediktor.

Untuk mengelompokkan tingkat pengangguran terbuka di Provinsi Jawa Timur, langkah-langkah analisis yang dilakukan adalah sebagai berikut.

1. Mengelompokkan wilayah untuk melihat wilayah-wilayah mana yang memiliki ciri khas yang sama.

Langkah-langkah yang dilakukan untuk memodelkan data dengan menggunakan *principal component regression* adalah sebagai berikut.

1. Melakukan standarisasi data pada variabel independen.
2. Membuat matrik korelasi antar variabel independen
3. Membangkitkan variabel baru PC yang saling independen

$$PC_1 = a_{11}Z_1 + a_{12}Z_2 + \dots + a_{1k}Z_k$$

$$PC_2 = a_{21}Z_1 + a_{22}Z_2 + \dots + a_{2k}Z_k$$

.....

$$PC_k = a_{k1}Z_1 + a_{k2}Z_2 + \dots + a_{kk}Z_k$$

4. Meregresikan variabel y dengan skor PC.
5. Memodelkan regresi Y dengan PC ke dalam model Y dengan z, kemudian x.

Langkah-langkah yang dilakukan untuk memodelkan data dengan menggunakan regresi ridge adalah sebagai berikut.

1. Melakukan standarisasi atau pembakuan data pada variabel dependen dan independen.
2. Menghitung Nilai koefisien $\hat{\beta}$ dengan berbagai nilai d .
3. Menentukan model yang stabil.
4. Melakukan uji signifikansi parameter regresi ridge.

Langkah-langkah yang dilakukan untuk memodelkan data dengan menggunakan analisis diskriminan adalah sebagai berikut.

1. Menguji homogenitas matriks varian kovarian dan normal multivariat.
2. Menguji kesamaan rata-rata kelompok.
3. Membentuk fungsi diskriminan berdasarkan kombinasi linier variabel prediktor.
4. Memilih fungsi diskriminan terbaik.
5. Menentukan ketepatan klasifikasi tingkat pengangguran terbuka Jawa Timur 2012 melalui nilai APER.

Halaman ini sengaja dikosongkan

BAB IV ANALISIS DAN PEMBAHASAN

Provinsi Jawa Timur merupakan provinsi dengan 29 kabupaten dan 9 kota yang memiliki beragam karakteristik pada masing-masing wilayah. Berdasarkan variabel-variabel yang mempengaruhi tingkat pengangguran terbuka (TPT) akan dilakukan deskripsi tentang wilayah pada kabupatn/kota di Jawa Timur Tahun 2012. Selanjutnya pengelompokan dan penanganan kasus multikolinieritas akan dilakukan dengan menggunakan *principal component regression* dan *ridge regression*, analisis *k-means cluster* serta analisis diskriminan.

4.1 Deskripsi Variabel Penelitian

Analisis deskriptif dilakukan untuk mengetahui karakteristik tingkat pengangguran terbuka (TPT) pada kabupaten/kota di Jawa Timur dan variabel-variabel prediktor terkait. Karakteristik yang dimaksud meliputi eksplorasi variabel respon dan prediktor sehingga diperoleh informasi yang lebih luas.

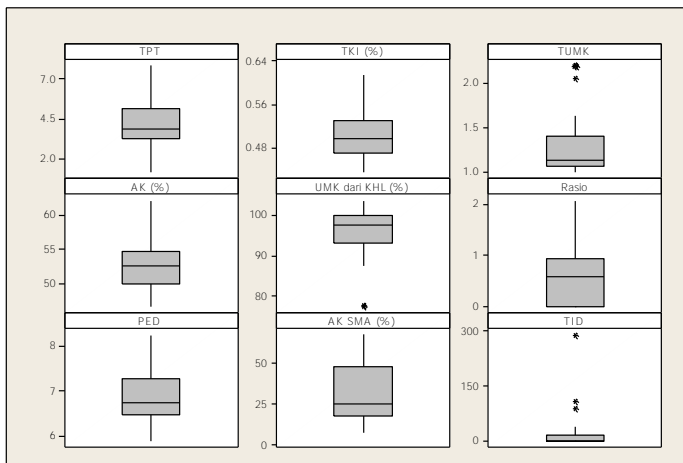
Tabel 4.1 Statistik Deskriptif Variabel Penelitian

Variabel	Rata-rata	Standar Deviasi	Minimum	Maksimum	Range
Y	4.230	1.698	1.160	7.850	6.690
X1	0.506	0.040	0.435	0.615	0.180
X2	1.315	0.376	1.000	2.200	1.200
X3	52.805	3.507	46.619	62.184	15.565
X4	96.101	5.799	77.452	103.783	26.331
X5	0.626	0.579	0.000	2.083	2.083
X6	6.908	0.539	5.890	8.250	2.360
X7	31.070	16.560	7.950	68.010	60.060
X8	18.600	51.270	0.000	291.190	291.190

Tabel 4.1 menunjukkan statistik deskriptif variabel-variabel penelitian meliputi secara umum. Statistik deskriptif yang dibahas dalam sub bab mencakup rata-rata, nilai minimum dan maksimum

hingga nilai range dan standar deviasi baik variabel respon maupun variabel prediktor.

Variabel prediktor yang memiliki range data terpanjang yaitu tingkat investasi daerah (X_8), dengan nilai minimum 0 dan nilai maksimum 291,19 sedangkan yang memiliki range data terpendek adalah variabel tingkat upah minimum (X_2) dengan nilai minimum 1,0 hingga 2,2. Nilai standar deviasi yang merupakan ukuran keragaman data menunjukkan bahwa variabel tingkat investasi daerah (X_8) memiliki keragaman data terbesar bila dibandingkan dengan variabel-variabel lainnya. Rata-rata persentase tenaga kerja yang bekerja ke luar negeri dari kabupaten/kota sebesar 0,506 persen atau dapat dikatakan rata-rata terdapat 5 tenaga kerja yang ke luar negeri dari 1000 angkatan kerja di masing-masing daerah. Di lain sisi, angkatan kerja yang berpendidikan SMA keatas (X_7) memiliki rata-rata sebesar 31,07 persen yang artinya rata-rata terdapat 310 angkatan kerja yang minimal berpendidikan SMA keatas dari 1000 angkatan kerja di masing-masing daerah.



Gambar 4.1 Boxplot TPT dan Faktor-Faktor yang Berpengaruh

Gambar 4.1 menunjukkan bahwa TPT dan variabel-variabel yang berpengaruh terlihat beragam pada beberapa kabupaten/kota di Jawa Timur. Pada beberapa variabel prediktor dapat diketahui

bahwa terdapat kabupaten/kota yang memiliki tanda bintang pada Gambar 4.1. Tanda bintang tersebut mengindikasikan terdapat *outlier* pada data.

4.2 Pemodelan Tingkat Pengangguran Terbuka

Dalam subbab ini akan dilakukan pemodelan pada tingkat pengangguran terbuka di Jawa Timur Tahun 2012 dengan variabel-variabel terkait. Beberapa metode yang digunakan antara lain analisis regresi linier berganda, deteksi kasus multikolinieritas, *principal component regression* dan regresi ridge.

4.2.1 Analisis Regresi Linier dan Deteksi Multikolinieritas

Analisis regresi pada sub bab ini akan menyatakan pola hubungan antara variabel TPT dengan variabel-variabel prediktor terkait. Persamaan regresi linier berganda pada kasus tingkat pengangguran terbuka ini dapat dinyatakan pada persamaan sebagai berikut.

$$Y = 2,84 - 188 x_1 - 0,079 x_2 + 18,3 x_3 - 0,00097 x_4 + 0,014 x_5 + 0,044 x_6 + 0,0019 x_7 + 0,0012 x_8$$

Tabel 4.2 ANOVA Tingkat Pengangguran Terbuka

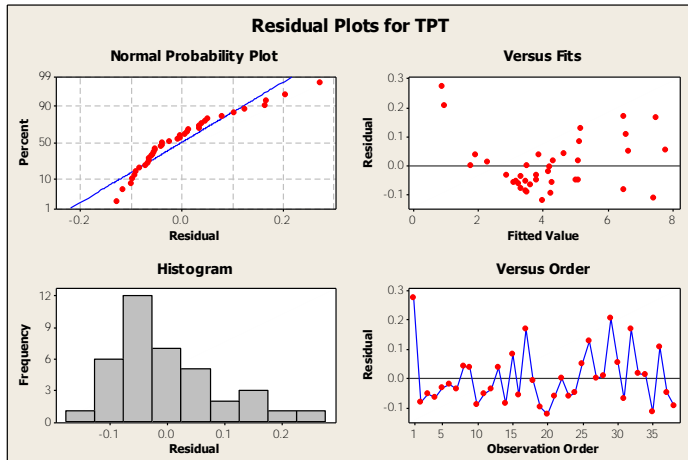
Source	DF	SS	MS	F	P
Regression	8	106,390	13,299	1190,980	0,000
Residual Error	29	0,324	0,011		
Total	37	106,714			

Tabel 4.2 menunjukkan bahwa bila dilakukan pengujian secara serentak persamaan regresi tersebut mempunyai pengaruh yang signifikan karena *p-value* kurang dari nilai *alpha* (0,05), jadi keputusan yang diambil adalah tolak H_0 . Di lain sisi, nilai *R-square* berada pada nilai sangat tinggi (99,7%) dan nilai *mean square error* yang kecil (0,0001). Dari hasil tersebut, model yang diperoleh bisa dikatakan bagus karena memiliki tingkat hubungan yang tinggi dan *mean square error* yang kecil. Namun bila ditinjau secara parsial, sebagian besar variabel prediktor menunjukkan hasil yang tidak signifikan karena *p-value* lebih

besar dari nilai *alpha*, seperti pada Tabel 4.3. Dari 8 variabel yang diregresikan, hanya terdapat 3 variabel saja yang signifikan. Oleh karena itu, akan dilakukan pengujian kolinieritas untuk mengetahui hubungan antar variabel prediktor.

Tabel 4.3 Uji Parsial Tingkat Pengangguran Terbuka

Prediktor	Koef	SE Koef	T	P	VIF
Konstan	2,840	0,496	5,720	0,000	
X ₁	-188,407	3,102	-60,740	0,000	51,863
X ₂	-0,079	0,075	-1,050	0,302	2,618
X ₃	1,828	0,034	53,990	0,000	46,723
X ₄	-0,001	0,004	-0,260	0,793	1,480
X ₅	0,014	0,033	0,420	0,677	1,185
X ₆	0,044	0,050	0,900	0,378	2,361
X ₇	0,002	0,002	1,300	0,202	2,098
X ₈	0,001	0,000	2,660	0,012	1,853



Gambar 4.2 Plot Residual Tingkat Pengangguran Terbuka

Di lain pihak, secara visual *normal probability plot* pada Gambar 4.2 menunjukkan terdapat beberapa plot residual yang lebih besar

(*outlier*), sehingga akan dilakukan pengujian *outlier* dengan menggunakan nilai DFFITS.

Gambar 4.2 menunjukkan terdapat beberapa pencilan data pada observasi ke 1 (kabupaten Pacitan), 25 (kabupaten Gresik) dan 29 (kabupaten Sumenep). Untuk mengetahui apakah ketiga data tersebut *outlier* ataukah tidak akan dilakukan pengujian dengan metode DFITS. Didapatkan hasil seperti pada Lampiran 7. Ketiga pencilan data yang ditunjukkan pada Gambar 4.2 memang mengindikasikan nilai *outlier* yang perlu dilakukan penanganan karena $|DFITS|$ lebih dari 0,97. Di lain sisi, kota Madiun juga teridentifikasi sebagai *outlier*. Maka terdapat 4 observasi yang perlu ditangani.

Tabel 4.4 Nilai Koefisien Korelasi Provinsi Jawa Timur

Variabel	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
X ₁	1	-0.429	0.986	-0.187	-0.288	-0.39	-0.573	-0.374
X ₂	-0.429	1	-0.409	0.323	0.18	0.582	0.365	0.612
X ₃	0.986	-0.409	1	-0.136	-0.264	-0.343	-0.515	-0.367
X ₄	-0.187	0.323	-0.136	1	-0.079	0.509	0.313	0.056
X ₅	-0.288	0.18	-0.264	-0.079	1	0.109	0.272	0.142
X ₆	-0.39	0.582	-0.343	0.509	0.109	1	0.564	0.176
X ₇	-0.573	0.365	-0.515	0.313	0.272	0.564	1	0.163
X ₈	-0.374	0.612	-0.367	0.056	0.142	0.176	0.163	1

Kasus multikolinieritas merupakan kasus dimana terjadi hubungan antara variabel prediktor. Bila ditinjau melalui nilai VIF (*Variance Inflation Factors*) pada Tabel 4.3, variabel persentase tenaga kerja indonesia (x_1) dan persentase angkatan kerja (x_2) masing-masing Kabupaten/Kota di Jawa Timur Tahun 2012 menunjukkan nilai 51,863 dan 46,723 ($VIF > 10$). Selain menggunakan nilai VIF, deteksi multikolinieritas juga dapat dilakukan dengan melihat korelasi antar variabel prediktor melalui koefisien korelasi. Pada Tabel 4.4 terlihat bahwa terdapat korelasi yang tinggi antara variabel persentase tenaga kerja Indonesia (x_1) dan persentase angkatan kerja masing-masing kabupaten (x_2), sehingga terdapat indikasi multikolinieritas pada kasus tersebut. Untuk menangani kasus multikolinieritas tersebut,

maka dilakukan penanganan dengan menggunakan metode *Principal Component Regression* (PCR) dan *Ridge Regresi*.

4.2.2 *Principal Component Regression* (PCR)

Principal Component Regression atau Regresi Komponen Utama merupakan salah satu metode yang biasa digunakan untuk mengatasi kasus multikolinieritas.

Tabel 4.5 *Eigen Value* Matrik Korelasi

Eigenvalue	3.579	1.355	1.076	0.850	0.512	0.391	0.224	0.010
Proportion	0.447	0.169	0.135	0.106	0.064	0.049	0.028	0.001
Cumulative	0.447	0.617	0.751	0.858	0.922	0.971	0.999	1

Lampiran 5 merupakan hasil dari transformasi data yang harus dilakukan terlebih dahulu. Tabel 4.5 dan Tabel 4.6 menunjukkan nilai *eigen value* dan *eigen vector* matrik korelasi. *Eigen value* pada Tabel 4.5 menunjukkan bahwa eigen value ≥ 1 , sehingga terdapat 3 variabel baru yang digunakan yaitu PC₁, PC₂ dan PC₃. Persamaan PC baru yang didapatkan bila meninjau Tabel 4.6 adalah

Tabel 4.6 *Eigen Vector* Matrik Korelasi

PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
-0.452	0.283	-0.187	-0.345	-0.143	0.128	-0.019	-0.724
0.391	0.146	-0.482	-0.179	-0.134	-0.295	-0.675	0.015
-0.433	0.320	-0.168	-0.389	-0.147	0.194	0.000	0.687
0.225	0.629	0.090	0.018	0.706	0.210	-0.035	-0.027
0.188	-0.406	0.195	-0.808	0.319	-0.061	0.059	-0.017
0.371	0.440	0.081	-0.185	-0.361	-0.429	0.560	-0.012
0.388	0.085	0.380	-0.104	-0.454	0.662	-0.201	-0.045
0.285	-0.184	-0.712	0.008	0.056	0.434	0.430	-0.015

$$PC_1 = -0.452 Z_1 + 0.391 Z_2 - 0.433 Z_3 + 0.225 Z_4 + 0.188 Z_5 + 0.371 Z_6 + 0.388 Z_7 + 0.285 Z_8$$

$$PC_2 = 0.283 Z1 + 0.146 Z2 + 0.320 Z3 + 0.629 Z4 - 0.406 Z5 + 0.440 Z6 + 0.085 Z7 - 0.184 Z8$$

$$PC_3 = -0.187 Z1 - 0.482 Z2 - 0.168 Z3 + 0.090 Z4 + 0.195 Z5 + 0.081 Z6 + 0.380 Z7 - 0.721 Z8$$

Pada Tabel 4.7 merupakan hasil regresi antara tingkat pengangguran terbuka dengan variabel PC_1 , PC_2 dan PC_3 . Variabel PC_2 dan PC_3 menunjukkan *p-value* lebih dari *alpha* (0,05) dengan nilai *R-square* dan *R-square (adj)* sebesar 62,7 persen dan 59,4 persen, sehingga keputusan yang diambil adalah gagal tolak H_0 maka variabel PC_2 dan PC_3 akan dihilangkan pada model regresi.

Tabel 4.7 Uji Parsial pada Variabel PC

Predictor	Coef	SE Coef	T	P	VIF
Constant	4.230	0.176	24.090	0.000	
PC1	0.674	0.093	7.270	0.000	1.000
PC2	-0.106	0.151	-0.700	0.486	1.000
PC3	0.328	0.169	1.940	0.061	1.000

Tabel 4.8 dan 4.9 menunjukkan baik secara pengujian serentak maupun secara parsial pada variabel PC_1 . Pada Tabel 4.7 didapatkan hasil yang signifikan dimana *p-value* kurang dari 0,05 (*alpha*) dengan *mean square error* sebesar 1,245. Di lain sisi, didapatkan nilai *R-square* dan *R-square (adj)* sebesar 58 persen dan 56,8 persen, sehingga 58 persen saja yang bisa dijelaskan melalui persamaan PC_1 dan sisanya dijelaskan oleh variabel lain. Nilai VIF pada Tabel 4.9 yang menunjukkan nilai kurang dari 10, sehingga tidak terdapat indikasi kasus multikolinieritas.

Tabel 4.8 ANOVA pada Variabel PC_1

Source	DF	SS	MS	F	P
Regression	1	61.881	61.881	49.690	0.000
Residual Error	36	44.834	1.245		
Total	37	106.714			

Tabel 4.9 Uji Parsial pada Variabel PC₁

Predictor	Coef	SE Coef	T	P	VIF
Constant	4.230	0.181	23.370	0.000	
PC1	0.674	0.096	7.050	0.000	1.000

Langkah selanjutnya adalah menentukan model regresi tersebut dalam variabel z dengan mensubstitusikan nilai PC₁

$$\hat{y} = 0.423 + 0.674 (0.452 Z_1 + 0.391 Z_2 - 0.433 Z_3 + 0.225 Z_4 + 0.188 Z_5) \\ + 0.371 Z_6 + 0.388 Z_7 + 0.285 Z_8)$$

selanjutnya menentukan model tersebut dalam variabel x

$$\hat{y} = 6.709 - 7.6496 x_1 + 0.7103 x_2 - 0.0844 x_3 + 0.0265 x_4 + 0.2214 x_5 + 0.47 x_6 \\ + 0.016 x_7 + 0.0038 x_8$$

Setiap variabel x_1 bertambah satu satuan maka tingkat pengangguran terbuka akan berkurang sebesar 7,65 satuan (dengan asumsi semua variabel konstan), namun bila ditinjau dari segi variabel-variabel lain seperti tingkat investasi daerah (x_8), persentase angkatan kerja berpendidikan SMA ketas (x_7) dan laju pertumbuhan ekonomi daerah (x_6) berbanding lurus dengan variabel respon tingkat pengangguran terbuka sehingga bila salah satu variabel ini bertambah sebesar satu satuan maka tingkat pengangguran terbuka juga akan naik sebesar koefisien pada masing-masing variabel.

4.2.3 Pengujian Asumsi Residual Identik

Salah satu asumsi yang penting model regresi adalah varians residual bersifat homoskedastisitas yaitu varians residual bersifat identik. Cara mendeteksi heteroskedastisitas adalah melalui uji *glejser* yang dilakukan dengan cara meregresikan harga mutlak residual $|e|$ dengan variabel prediktor, yang dalam hal ini adalah PC₁.

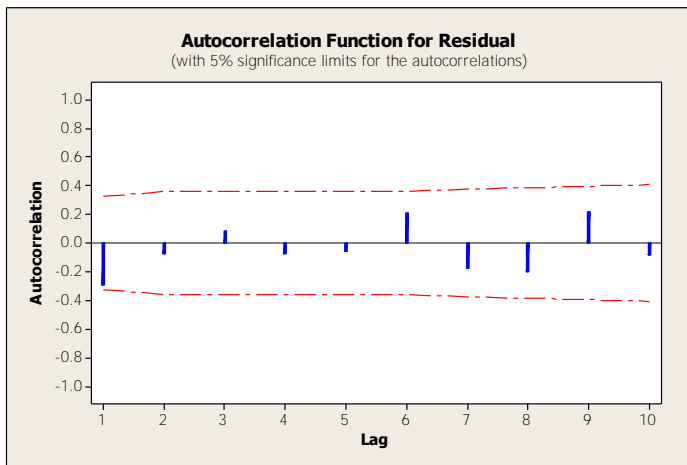
Tabel 4.10 Uji Parsial pada Uji *Glejser*

Predictor	Coef	SE Coef	T	P
Constant	0.000	0.181	0.000	1.000
PC1	0.000	0.096	0.000	1.000

Pada Tabel 4.10 dapat diketahui bahwa variabel PC_1 tidak signifikan, maka kesimpulan yang diambil adalah gagal tolak H_0 ($p\text{-value} > 0,05$) yang berarti residual cenderung identik. Maka asumsi residual identik telah terpenuhi.

4.2.4 Pengujian Asumsi Residual Independen

Salah satu metode yang dapat digunakan untuk menguji independensi adalah dengan menggunakan plot ACF (*Autocorrelation Function*) pada residual. Dan diperoleh plot ACF pada Gambar 4.3

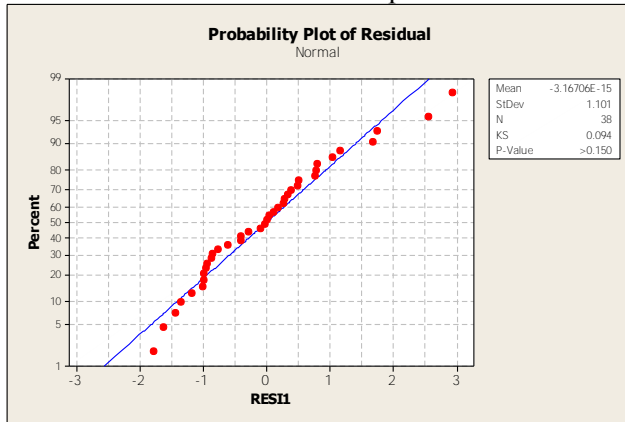


Gambar 4.3 Plot ACF pada Residual

karena pada Gambar 4.3 tidak terdapat nilai lag yang keluar dari batas signifikansi, maka keputusan yang diambil adalah gagal tolak H_0 . Plot *autocorrelation function* menunjukkan residual yang independen.

4.2.5 Pengujian Asumsi Residual Berdistribusi Normal

Pengujian asumsi residual Normal ($0, \sigma^2$) dapat dilakukan melalui uji Kolmogorov Smirnov. Secara visual Gambar 4.4 menunjukkan bahwa plot-plot residual terlihat mendekati garis linier yang mengindikasikan asumsi residual berdistribusi normal terpenuhi. Dan hal tersebut diperkuat oleh p -value yang menunjukkan nilai lebih besar dari α maka mengindikasikan bahwa residual berdistribusi normal terpenuhi.



Gambar 4.4 Plot Residual Distribusi Normal

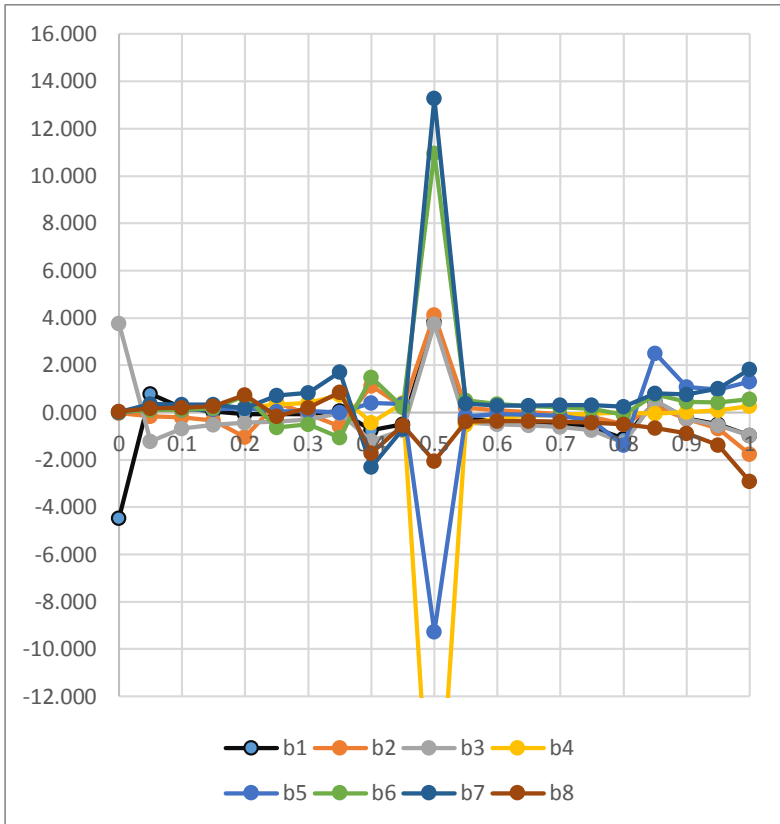
Pengujian asumsi IIDN telah terpenuhi, namun karena nilai R -square yang cukup kecil, maka akan dilakukan pemodelan menggunakan regresi ridge sehingga diperoleh hasil yang lebih stabil.

4.2.6 Analisis Ridge Regression pada TPT

Regresi ridge digunakan untuk mengulangi masalah multikolinieritas adalah dengan cara menambahkan suatu nilai d yang bernialai antara $0 < d < 1$ pada diagonal matriks $\mathbf{Z}^T\mathbf{Z}$. Nilai koefisien $\hat{\beta}$ untuk berbagai tetapan bias d untuk pemodelan tingkat pengangguran terbuka Provinsi Jawa Timur dapat dilihat pada Tabel 4.11.

Tabel 4.11 Nilai $\hat{\beta}$ dengan berbagai Nilai d

d	Z₁	Z₂	Z₃	Z₄	Z₅	Z₆	Z₇	Z₈
0	-4.475	-0.017	3.775	-0.003	0.005	0.014	0.019	0.037
0.05	0.792	-0.162	-1.224	0.208	0.130	0.107	0.362	0.168
0.1	0.206	-0.210	-0.675	0.199	0.121	0.115	0.339	0.195
0.15	0.034	-0.336	-0.516	0.209	0.128	0.173	0.333	0.278
0.2	-0.064	-1.039	-0.436	0.202	0.185	0.687	0.183	0.735
0.25	-0.062	0.397	-0.377	0.328	0.051	-0.622	0.718	-0.157
0.3	-0.075	-0.062	-0.300	0.406	0.071	-0.490	0.826	0.202
0.35	0.063	-0.582	-0.025	0.745	0.008	-1.054	1.722	0.864
0.4	-0.753	1.121	-1.204	-0.424	0.414	1.495	-2.300	-1.717
0.45	-0.495	0.248	-0.748	0.406	0.369	0.232	-0.707	-0.542
0.5	3.829	4.123	3.745	-20.794	-9.266	10.949	13.274	-2.056
0.55	-0.267	0.206	-0.441	-0.511	-0.161	0.521	0.382	-0.390
0.6	-0.343	0.096	-0.500	-0.238	-0.076	0.353	0.296	-0.353
0.65	-0.400	0.018	-0.545	-0.138	-0.083	0.279	0.296	-0.353
0.7	-0.473	-0.063	-0.610	-0.082	-0.148	0.224	0.312	-0.377
0.75	-0.604	-0.178	-0.742	-0.042	-0.346	0.153	0.324	-0.423
0.8	-1.096	-0.492	-1.282	0.010	-1.368	-0.097	0.245	-0.493
0.85	0.478	0.244	0.511	-0.049	2.496	0.774	0.821	-0.663
0.9	-0.239	-0.264	-0.275	0.017	1.084	0.446	0.766	-0.883
0.95	-0.501	-0.671	-0.533	0.082	0.970	0.427	1.005	-1.370
1	-0.975	-1.756	-0.966	0.269	1.303	0.562	1.824	-2.909



Gambar 4.5 Grafik Nilai $\hat{\beta}$ dengan berbagai Nilai d

Berdasarkan Tabel 4.11 dan Gambar 4.5 diketahui bahwa dengan nilai $d = 0,8$ terlihat bahwa koefisien $\hat{\beta}$ lebih stabil. Nilai d yang diperoleh digunakan untuk memodelkan regresi ridge sehingga model regresi yang terbentuk adalah sebagai berikut

$$\hat{y} = -1,096 Z_1 - 0,492 Z_2 - 1,282 Z_3 + 0,009 Z_4 - 1,368 Z_5 + \\ - 0,097 Z_6 + 0,245 Z_7 - 0,493 Z_8$$

Berdasarkan model regresi ridge untuk mengetahui signifikansi dari koefisien masing-masing variabel prediktor dilakukan uji koefisiensi regresi ridge dengan hipotesis

$H_0: \beta_k = 0$ (koefisien regresi tidak signifikan)

$H_1: \beta_k \neq 0$ (koefisien regresi signifikan)

Apabila $|t_{hit}| > t_{(\alpha/2; p; n-p-1)} = 2,045$, maka kesimpulan yang diambil adalah Tolak H_0 .

Tabel 4.12 Uji Signifikansi Parameter Regresi Ridge

β_k	Koefisien	Varsians	t_{hit}	$ t_{hit} $
β_1	-1.0962	0.0016270	-83.422	83.422*
β_2	-0.4916	0.0014153	-30.274	30.274*
β_3	-1.2818	0.0012299	-105.966	105.966*
β_4	0.0095	0.0033623	0.344	0.344
β_5	-1.3679	0.0033477	-47.995	47.995*
β_6	-0.0974	0.0029027	-4.113	4.113*
β_7	0.2451	0.0274359	3.261	3.261*
β_8	-0.4933	0.2628591	-2.076	2.076*

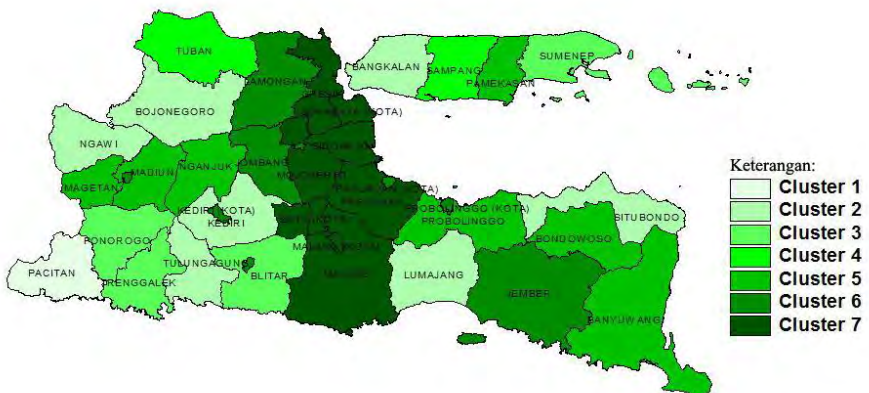
Keterangan: *) Uji signifikansi parameter model regresi signifikan dengan α 5%

Tabel 4.12 menunjukkan koefisien parameter model regresi dan nilai t hitung yang digunakan untuk menguji signifikansi parameter regresi ridge dengan cara membandingkan nilai statistik uji t dengan $t_{(0.05; 29)} = 2,045$. Berdasarkan uji signifikansi parameter regresi ridge didapatkan kesimpulan bahwa hampir semua parameter telah signifikan artinya variabel-variabel yang mempengaruhi tingkat pengangguran terbuka Provinsi Jawa Timur adalah persentase tenaga kerja Indonesia (x_1), tingkat UMK (x_2), persentase angkatan kerja Indonesia (x_3) dan rasio banyak perusahaan (x_5). Setiap penambahan 1 persen tenaga kerja Indoensia maka TPT di Jawa Timur akan berkurang sebesar 1,096 persen. Di sisi lain, setiap angkatan kerja bertambah 1 persen maka TPT cenderung bertambah sebesar 1,282 persen, sedangkan setiap rasio banyaknya perusahaan per 100 angkatan kerja

bertambah 1 satuan maka TPT cenderung berkurang sebesar 1,368 satuan.

4.3 Analisis *K-Means Cluster*

K-means cluster merupakan salah satu pengelompokan bukan hirarki yang bertujuan mengelompokkan n objek (kabupaten/kota) ke dalam k kelompok. Dalam kasus tingkat pengangguran terbuka ini akan dikelompokkan menjadi 7 kelompok berdasarkan variabel-variabel prediktor yang digunakan.



Gambar 4.6 Peta *K-means Cluster* di Jawa Timur Tahun 2012

Pemetaan hasil pengelompokan dengan menggunakan *k-means cluster* dapat dilihat pada Gambar 4.6. Letak kabupaten/kota yang tergabung dalam satu *cluster* (kelompok) saling berdekatan. Meski begitu, ada kabupaten/kota yang letaknya berjauhan. Setiap kelompok yang dihasilkan dengan suatu metode pengelompokan tentu memiliki karakteristik unik yang berbeda satu sama lainnya. *Cluster 1* mempunyai persentase tenaga kerja Indonesia dan angkatan kerja yang paling tinggi. *Cluster 2* cenderung memiliki laju pertumbuhan daerah di bawah rata-rata.

Sementara itu, *cluster 3* adalah kumpulan kabupaten yang memiliki tingkat upah minimum paling rendah bila dibandingkan

dengan kabupaten/kota lain. *Cluster* 3 juga memiliki rasio perusahaan per 100 angkatan kerja yang sedikit. Di lain pihak, *cluster* 4 merupakan kelompok yang memiliki persentase upah minimum yang paling rendah bila dibandingkan dengan kondisi hidup layak masing-masing daerah. *Cluster* 5 merupakan kelompok daerah yang tergolong pada kelas menengah hampir pada setiap variabel prediktor, mencakup variabel tenaga kerja Indonesia, tingkat upah minimum, angkatan kerja hingga laju pertumbuhan ekonomi daerah. *Cluster* 6 mempunyai keunggulan pada laju pertumbuhan ekonomi daerah yang paling tinggi bila dibandingkan dengan daerah-dearah lain. Dan hampir setiap kota seperti kota Kediri, Blitar, Mojokerto hingga Madiun termasuk pada kelompok 6. *Cluster* 7 dengan presentase tenaga kerja Indonesia dan angkatan kerja yang cenderung rendah memiliki upah minimum dari kondisi hidup layak yang tinggi. Di lain pihak, *cluster* 7 juga memiliki banyak perusahaan dan tingkat investasi daerah yang tinggi serta laju pertumbuhan ekonomi daerah di atas rata-rata. Kelebihan dan kekurangan yang dimiliki setiap kelompok seharusnya bisa dikembangkan dan segera diberi perhatian khusus.

4.4 Analisis Diskriminan Tingkat Pengangguran Terbuka

Analisis diskriminan merupakan metode yang digunakan untuk mengetahui perbedaan kelompok dari metode *cluster* berdasarkan variabel-variabel yang mempengaruhi tingkat pengangguran terbuka di Provinsi Jawa Timur. Dalam analisis diskriminan, beberapa asumsi yang harus terpenuhi terlebih dahulu adalah asumsi homogenitas matriks varian kovarian dan normal multivariat.

Tabel 4.13 Pengujian Asumsi pada Analisis Diskriminan

Normal Multivariat	0.579
Box's M	0.088

Pengujian kehomogenan matriks varian kovarian menggunakan Box's *M Test* dengan hipotesis

$$H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5$$

H_1 : minimal ada satu Σ_j yang berbeda

Hasil pengujian dengan menggunakan α sebesar 0,05 dapat disimpulkan bahwa data matriks varians kovarians sama, karena nilai *p-value* dari Box's *M Test* lebih dari *alpha* (0,05). Keputusan yang diambil adalah gagal tolak H_0 , sehingga asumsi homogenitas matriks varians kovarian terpenuhi.

Asumsi normal multivariat pada variabel prediktor dapat diketahui melalui statistik uji dengan hipotesis

H_0 : data berdistribusi multivariat normal

H_1 : data tidak berdistribusi multivariat normal

karena nilai *p-value* dari normal multivariat lebih dari 0,5, maka keputusan yang diambil adalah gagal tolak H_0 . Asumsi data berdistribusi normal multivariat terpenuhi. Maka asumsi homogenitas matriks varian kovarian dan normal multivariat telah terpenuhi.

Tabel 4.14 Uji Kesamaan Rata-Rata Kelompok

Variabel	Sig.
X ₁	.000
X ₂	.000
X ₃	.000
X ₄	.000
X ₅	.023
X ₆	.000
X ₇	.000
X ₈	.142

Berdasarkan Tabel 4.14 tentang uji kesamaan rata-rata kelompok, terlihat mayoritas variabel memiliki *p-value* lebih kecil dari pada *alpha* (0,05), sehingga mayoritas variabel memiliki perbedaan rata-rata.

Fungsi diskriminan dibentuk berdasarkan pada kombinasi linear dari variabel-variabel prediktor yang memberikan pembeda terbaik. Variabel yang berperan nyata dalam membedakan kelompok-kelompok tersebut adalah persentase tenaga kerja

Indonesia (X_1), tingkat upah minimum (X_2), persentase upah minimum dari kondisi hidup layak (X_4) dan laju pertumbuhan ekonomi daerah (X_6). Maka fungsi diskriminan terbaik untuk membedakan kelompok kabupaten/kota di Provinsi Jawa Timur adalah

$$\text{Fungsi 1} = -33,21 - 7,13 X_1 + 4,88 X_2 + 0,14 X_4 + 2,43 X_6$$

$$\text{Fungsi 2} = -40,32 + 43,36 X_1 + 0,78 X_2 + 0,19 X_4 - 0,095 X_6$$

$$\text{Fungsi 3} = 5,49 + 17,72 X_1 + 3,45 X_2 - 0,18 X_4 - 0,28 X_6$$

$$\text{Fungsi 4} = -7,96 + 7,48 X_1 - 1,72 X_2 - 0,11 X_4 + 2,41 X_6$$

Tabel 4.15 *Eigen Value dan Canonical Correlation*

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	12.787	74.1	74.1	.963
2	3.031	17.6	91.7	.867
3	1.282	7.4	99.1	.749
4	.157	.9	100.0	.368

Nilai *canonical correlation* pada Tabel 4.15 menunjukkan bahwa pada fungsi 1 diperoleh nilai 0,963 atau dengan kata lain model mampu menjelaskan 96,3 persen keragaman dari variabel respon. Model pada fungsi 2 mampu menjelaskan 86,7 persen keragaman dari variabel respon, sedangkan model pada fungsi 4 hanya mampu menjelaskan 36,8 persen keragaman dari variabel respon saja.

Persamaan fungsi diskriminan linier yang terbentuk untuk setiap kelompok yang akan membedakan setiap *cluster* berdasarkan variabel-variabel yang mempengaruhi tingkat pengangguran terbuka pada kabupaten/kota di Provinsi Jawa Timur dapat dilihat pada Tabel 4.16. Fungsi diskriminan secara umum dapat dievaluasi ketepatan modelnya melalui *classification result (original)*, Tabel 4.17.

Tabel 4.16 Koefisien Fungsi Kelompok

Vaariabel	Cluster						
	1	2	3	4	5	6	7
X_1	1713.3	1457.2	1602.3	1418.2	1519.7	1391.0	1459.6
X_2	152.8	149.6	149.7	145.0	152.8	157.6	191.4
X_4	12.7	12.2	12.4	10.4	12.2	12.3	13.1
X_6	93.1	92.0	91.4	85.6	93.4	100.5	111.4
(Constant)	-1534.7	-1337.4	-1425.9	-1115.4	-1383.9	-1380.3	-1627.4

Evaluasi fungsi klasifikasi berguna untuk menentukan besarnya nilai kesalahan klasifikasi kabupaten/kota di Provinsi Jawa Timur.

Tabel 4.17 *Classification Result (Original)*

	CLUSTER	Predicted Group Membership						
		1	2	3	4	5	6	7
Original	1	1	0	0	0	0	0	0
	2	0	5	0	0	2	0	0
	3	0	0	4	0	0	0	0
	4	0	0	0	2	0	0	0
	5	0	1	1	0	5	0	0
	6	0	1	0	0	0	8	0
	7	0	0	0	0	0	0	8

Pada Tabel 4.17, terlihat bahwa nilai *Apperent Error Rates* (APER) untuk hasil klasifikasi kabupaten/kota di Jawa Timur berdasarkan faktor-faktor yang mempengaruhi tingkat pengangguran terbuka sebesar 0,132 yang berarti bahwa akurasi pada model mencapai 86,8 persen.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan dapat diambil kesimpulan sebagai berikut.

1. Karakteristik kabupaten/kota di Provinsi Jawa Timur yaitu rata-rata terdapat 4,23 persen tingkat pengangguran terbuka dengan keragaman data variabel tingkat investasi daerah (X_8) yang paling besar bila dibandingkan dengan variabel-variabel lainnya. Rata-rata persentase tenaga kerja yang bekerja ke luar negeri dari kabupaten/kota sebesar 0,506 persen atau dapat dikatakan rata-rata terdapat 5 tenaga kerja yang ke luar negeri dari 1000 angkatan kerja di masing-masing daerah. Di lain sisi, angkatan kerja yang berpendidikan SMA keatas (X_7) memiliki rata-rata sebesar 31,07 persen yang artinya rata-rata terdapat 310 angkatan kerja yang minimal berpendidikan SMA keatas dari 1000 angkatan kerja di masing-masing daerah.
2. Pemodelan tingkat pengangguran terbuka di Provinsi Jawa Timur berdasarkan variabel berpengaruh adalah
$$\hat{y} = -1,096 Z_1 - 0,492 Z_2 - 1,282 Z_3 + 0,009 Z_4 - 1,368 Z_5 + - 0,097 Z_6 + 0,245 Z_7 - 0,493 Z_8$$
3. Pengelompokan kasus tingkat pengangguran terbuka kabupaten/kota di Jawa Timur dibagi menjadi 7 kelompok (*cluster*) menggunakan metode *k-means*. *Cluster* 1 mempunyai persentase tenaga kerja Indonesia dan angkatan kerja yang paling tinggi. *cluster* 3 adalah kumpulan kabupaten yang memiliki tingkat upah minimum dan rasio perusahaan paling rendah. *Cluster* 7 dengan presentase tenaga kerja Indonesia dan angkatan kerja yang cenderung rendah memiliki upah minimum dari kondisi hidup layak yang tinggi. Di lain pihak, *cluster* 7 juga memiliki banyak perusahaan dan tingkat investasi daerah yang tinggi serta laju pertumbuhan ekonomi daerah di atas rata-rata.

4. Nilai *Apperent Error Rates* (APER) klasifikasi kabupaten/kota di Jawa Timur berdasarkan faktor-faktor yang mempengaruhi tingkat pengangguran terbuka sebesar 0,132 yang berarti bahwa akurasi pada model mencapai 86,8 persen.

5.2 Saran

Upaya yang sebaiknya dilakukan pemerintah adalah menekan tingkat pengangguran terbuka di Jawa Timur dengan membuka lapangan kerja yang sebanyak-banyaknya. Di sisi lain, kasus tentang tingkat pengangguran terbuka ini mengindikasikan pola non linier dan perlu dilakukan pengujian lebih lanjut tentang hal tersebut.

BIODATA PENULIS



Penulis yang bernama lengkap Nugroho Manggala Putra dan akrab disapa Boy dilahirkan pada tanggal 6 Juli 1992 di Mojokerto. Pendidikan formal yang telah ditempuh oleh penulis sebelum memasuki dunia perkuliahan yaitu SD Negeri Gambiran 1 Mojoagung (1998), SMP Negeri 1 Mojoagung (2004) dan SMA Negeri Mojoagung (2007). Tahun 2010 penulis diterima sebagai Mahasiswa

Baru S1 jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember Surabaya dengan nomer induk mahasiswa NRP 1310 100 019. Selama perkuliahan penulis lebih suka terjun ke dalam proyek-proyek lapangan seperti survei maupun magang di beberapa perusahaan. Apabila pembaca ingin berdiskusi mengenai tugas akhir ini, penulis dapat dihubungi melalui email:

nugroho.manggala.p@gmail.com